MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 710–725 doi: 10.1111/mec.13055

INVITED REVIEWS AND SYNTHESES

Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?

MARIANO ALVAREZ,* AARON W. SCHREY† and CHRISTINA L. RICHARDS*

Department of Integrative Biology, University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620, USA, †Department of Biology, Science Center, Armstrong State University, 11935 Abercorn Street, Savannah, GA 31419, USA

Abstract

Molecular ecology has moved beyond the use of a relatively small number of markers, often noncoding, and it is now possible to use whole-genome measures of gene expression with microarrays and RNAseq (i.e. transcriptomics) to capture molecular response to environmental challenges. While transcriptome studies are shedding light on the mechanistic basis of traits as complex as personality or physiological response to catastrophic events, these approaches are still challenging because of the required technical expertise, difficulties with analysis and cost. Still, we found that in the last 10 years, 575 studies used microarrays or RNAseq in ecology. These studies broadly address three questions that reflect the progression of the field: (i) How much variation in gene expression is there and how is it structured? (ii) How do environmental stimuli affect gene expression? (iii) How does gene expression affect phenotype? We discuss technical aspects of RNAseq and microarray technology, and a framework that leverages the advantages of both. Further, we highlight future directions of research, particularly related to moving beyond correlation and the development of additional annotation resources. Measuring gene expression across an array of taxa in ecological settings promises to enrich our understanding of ecology and genome function.

Keywords: ecological transcriptome, gene expression, microarrays, RNAseq, transcriptomics

Received 31 March 2014; revision received 16 December 2014; accepted 18 December 2014

Introduction

The interactions between organisms and environments are of central importance to many questions in the study of ecology (Scheiner & Willig 2011). While much progress has been made by examining traits and behaviours of individuals within and among populations, the integration of molecular techniques into ecology has allowed investigators unprecedented ability to assess the mechanisms that govern ecological interactions and underlie pattern and process at the most basic levels of biological organization. Now, genomewide approaches can lay the foundation for sophisticated functional studies that explore the specific genomic basis of phenotypic variation and rapid response to environmental change (e.g. mass flowering in a tropical tree Kobayashi *et al.* 2013; plasticity underlying response to osmotic conditions in killifish, Whitehead *et al.* 2013). As a result, the past decade has seen the rise of molecular ecology as a synthetic discipline that uses molecular techniques to answer (and often generate) ecological questions (Andrew *et al.* 2013). DNA microarrays and RNA sequencing (RNAseq) are the most widespread and powerful transcriptomics technologies, and allow ecologists to simultaneously measure genomewide gene expression on large numbers of individuals in wild populations. These tools measure variation in gene expression at the level of mRNA, which ultimately contributes to the formation of proteins, cellular phenotype and organismal phenotype that can be shaped by ecological processes (Oleksiak *et al.* 2002; Crawford & Oleksiak 2007; Aubin-Horth & Renn 2009; Ekblom & Galindo 2011; Whitehead 2012). The quantification of these patterns on a genomewide scale allows us to observe the molecular regulation of phenotype in response to ecological phenomena, and begin to understand the ecological transcriptome (Richards *et al.* 2009).

Correspondence: Christina L. Richards, Fax: 813-974-3263; E-mail: clr@usf.edu

Application of transcriptomics in an ecological context has become imperative because as the research community continues to acquire abundant genomics data for a variety of organisms in controlled laboratory settings, we have made little progress in understanding how the genome actually functions to create complex traits and adapt to complex environments (Richards *et al.* 2009). The nascent field of ecological genomics has already begun to shed light on how genomes function in natural environments including the mechanisms underlying adaptation (e.g. Lai *et al.* 2006; Elmer & Meyer 2011; Andrew *et al.* 2013), divergence (e.g. Pavey *et al.* 2010; Renaut & Bernatchez 2011; Nosil & Feder 2012), genotype-by-environment interactions (e.g. Richards *et al.* 2012) and phenotypic plasticity (e.g. Wittkopp 2007; Aubin-Horth & Renn 2009; Nicotra *et al.* 2010; Whitehead *et al.* 2013; Schneider *et al.* 2014). These studies capitalize on the statistical strength of ecological experimental design to capture sources of variation while leveraging powerful genomics tools to assess gene activity. The work in *Helianthus* by Rieseberg *et al.* serves as an illustrative example. With classic ecological design in several greenhouse and field studies, they documented that some adaptive traits in wild populations of the hybrid sunflower species *Helianthus deserticola* are much greater or much smaller (i.e. transgressive) compared to the parental species *Helianthus annus* and *Helianthus petiolaris* (Rosenthal *et al.* 2002; Gross *et al.* 2004). Lai *et al.* (2006) used microarray data from all three species to suggest that novel gene expression in the hybrid *H. deserticola* may contribute to the transgressive phenotypic patterns. In particular, they identified a number of highly differentially transcribed transporter genes and speculated that the differential expression of these genes was correlated with adaptation in *H. deserticola* to an extreme, arid environment. This was supported by an association between fitness and expression of one of the genes of interest (G protein-coupled receptor: QHB30N12) in *H. deserticola* in the field. The combination of molecular techniques with a traditional ecological design allowed Lai *et al.* (2006) to identify possible mechanisms of adaptation that resulted from hybridization. This union of ecology and molecular biology is the hallmark of molecular ecology.

Why wild systems?

By wild systems, we typically mean nonmodel organisms or nontraditional model organisms (e.g. *Daphnia pulex*, *Coregonus clupeaformis*, *Fundulus heteroclitus*) in natural settings, although traditional model organisms like *Arabidopsis thaliana*, *Drosophila melanogaster* and *Caenorhabditis elegans* do have their own ecology and can be extremely informative when studied in an ecological setting (Kammenga *et al.* 2007; Pavey *et al.* 2012; Weigel 2012). As organisms in natural settings are continuously exposed to multiple environmental signals and must respond appropriately to dynamic conditions, this context provides a unique opportunity to discover information about gene expression patterns that cannot be gleaned through controlled laboratory settings. Transcriptome studies in natural settings have found novel expression of otherwise-silent genes that only respond to the multiple, simultaneous stimuli that occur in complex, dynamic environments (Colbourne *et al.* 2011; Pavey *et al.* 2012). Novel behaviour of transcripts may also be exposed by particular environmental challenges, which may contribute to variation between individuals and populations (Dalziel *et al.* 2009). For example, Whitehead *et al.* (2012) examined pollution-tolerant and naïve populations of *F. heteroclitus* along the northern Atlantic coast of the United States. Divergence in expression of genes that were responsive to the toxin was revealed only at higher dosages of PCB-126. On the other hand, the authors found that neutral processes explained patterns of population divergence in expression of genes that were not responsive to dioxin-like compound PCB-126. These results suggest that environmental challenge may be necessary to expose adaptive population divergence: without the stimulus of high PCB concentrations, the population differences between *F. heteroclitus* were obscured (Whitehead *et al.* 2012). Further, the complex interactions of environmental factors in natural systems may reveal more differentiation between populations than would be observable under controlled conditions.

Besides identifying context dependency of transcription, transcriptome studies in nonmodel organisms may yield functional information about novel transcripts that either have no homolog in their most closely related model organism, or have taken on a novel function. In *Daphnia pulex*, researchers found that unannotated, *Daphnia*-specific genes made up more than a third of the transcriptome and were the most responsive to a variety of ecological stimuli (Colbourne *et al.* 2011). The study further revealed the importance of the diversification of duplicated genes within specific metabolic pathways in this species. In many cases, gene duplication allowed for immediate divergence in expression patterns, which may be particularly effective when the duplicated gene interacts with genes sharing a common regulatory program. This novel behaviour or function of duplicated genes was only exposed under specific environmental challenges, suggesting that regardless of the number of laboratory studies on an organism, a large number of gene functions cannot be annotated without exposure to complex natural stimuli.

Box 1. Primary problems in transcriptome studies

Bias in signals

DNA microarray and RNAseq data each display biases and distortions at different ends of the gene expression spectrum. RNAseq is biased towards highly transcribed genes (Łabaj *et al.* 2011; Malone & Oliver 2011). The genes that are more highly transcribed have more abundant transcripts and are more likely to be sequenced, leaving less highly transcribed genes with comparatively less sequencing coverage (Łabaj *et al.* 2011), which potentially reduces the resolution of more subtle patterns of gene expression. Rather than directly counting transcripts, microarrays depend on fluorescently labelled targets that hybridize to probes. Each gene has a measurement device (the probes) that saturates at high expression levels, but the probes will detect genes that are expressed at lower levels. Thus, microarrays may be a more appropriate choice for the detection of genes that are expressed in low-abundance (Łabaj *et al.* 2011; Malone & Oliver 2011). However, due to the fluorescence-based quantification method, microarrays experience some compression at the higher end of expression. This reduces the ability of microarrays to quantify very highly expressed genes (Malone & Oliver 2011). A user should carefully consider which bias is more tolerable for the ecological question.

Heterologous arrays

This problem only relates to microarray studies where microarrays of related species are used to characterize gene expression in a species with no genomic resources (the so-called heterologous arrays). While heterologous arrays can be useful, they must be used with caution because of mishybridization between probes designed for one species and RNA extracted from a different species (Buckley 2007). The problem of probe mishybridization was made famous by a comparison that found that gene expression differences in human brains were much greater than those in any chimpanzee tissues (Enard *et al.* 2002). However, Hsieh *et al.* (2003) showed that the use of short-oligonucleotide microarrays biased the results because some of the probes did not hybridize efficiently to the chimpanzee cDNA (Buckley 2007). A reanalysis with long oligonucleotide arrays, which are less sensitive to polymorphisms, revealed that the patterns in expression from the brain tissue were actually less divergent (Hsieh *et al.* 2003). In addition to highlighting some of the potential problems with microarray technology, this study provided a first glance at the complexity involved in studying divergence between two closely related species.

Polyploidy

Polyploid organisms may express many duplicate genes, and RNAseq-based transcriptomes that are assembled de novo (as nonmodel organisms often are) may align transcripts from different gene copies that have different function (Ilut *et al.* 2012). Further, when a polyploid organism is compared to a diploid reference genome, transcripts from duplicated genes may confound the relative expression of those genes (Ilut *et al.* 2012). Both of these issues may cause errors when inferring gene expression. Because RNAseq data provide information about polymorphisms, newer bioinformatics pipelines for sequence data, such as PolyCat (Page *et al.* 2013) and HomeoRoq (Akama *et al.* 2014), may alleviate these issues. PolyCat, developed for cotton, uses SNP information from related diploid species to accurately map sequencing reads from coresident genomes of allopolyploids (Page *et al.* 2013). HomeoRoq was developed specifically for RNAseq data and uses parental genomes to identify the ratio of gene expression from coresident allopolyploid genomes (Akama *et al.* 2014). These methods are some of the first attempts to decipher the relative contributions of duplicate genes in allopolyploids using genomewide data. Because microarrays rely on hybridization and not direct sequencing, they provide a biologically relevant readout of the amount of gene transcript regardless of how many copies are contributing (with carefully designed probes). However, because similar sequences may cohybridize with the same probe, microarrays are unable to discriminate between duplicated or highly similar genes, and they cannot describe the relative contributions of the hybridized transcripts.

RNA pooling

Pooling RNA samples from multiple individuals before cDNA conversion allows multiple individuals to be screened on the same microarray or sequencing lane; therefore, population representation is increased without increasing cost (Pronk *et al.* 2011; e.g. Zhang & Gant 2005; Richards *et al.* 2012). The utility of pooling relies on the concept of biological averaging, pooled transcript abundance represents an average of the expression states among pooled samples (Kendziorski *et al.* 2005; Zhang & Gant 2005). However, pooling introduces a number of artefacts into the data. First, overall expression variability is reduced (Kendziorski *et al.* 2005). While this may be advantageous for field studies that can be extremely variable, it may lower precision of detection of expression levels of some genes (Kendziorski *et al.* 2005). Second, the measured expression will be more attenuated. In a pooled design, genes are averaged twice—once biologically because of pooling, and once technically during data normalization, which results in nonlinear distortion (Kendziorski *et al.* 2005). Third, genes that are expressed at a lower level are more affected by the distortion introduced by pooling (Pronk *et al.* 2011), and differential expression of these genes may go undetected. Thus, RNA pooling may exacerbate the problem of minimally expressed genes having larger effects on phenotype despite a small change in abundance (Oleksiak *et al.* 2004). Finally, because it is not possible to separate individuals from an RNA pool, pooling results in the loss of the ability to measure individual differences.

Statistical analysis

Classic analysis of variance (ANOVA) has been adapted to the interpretation of gene expression and gene-specific modelling of microarray data by fitting a global normalization model incorporating all of the genes, and then running a separate ANOVA for each gene (Wolfinger *et al.* 2001; Ayroles & Gibson 2006). The analysis of RNAseq data is not as mature and a consensus does not yet exist for RNAseq on preprocessing, normalization and inference methods. RNAseq data are generally described as an overdispersed Poisson distribution (Kvam *et al.* 2012; Wolf 2013), so familiar analyses and software that rely on normally distributed data, like the R package limma, may not be suitable for RNAseq data (Kvam *et al.* 2012). Law *et al.* (2014) have proposed a newer methodology, which generates a precision weight for each observation. This weighting system allows users to analyse RNAseq data as normally distributed data, which would allow analysis to follow the methods previously described for microarrays (Law *et al.* 2014). However, RNAseq facilitates analysis methodologies not available for microarray data. RNAseq detects transcript polymorphisms (Ekblom & Galindo 2011), which allows for the investigation of molecular evolution (Williams & Oleksiak 2008) and population genetics (Williams *et al.* 2010). This capability opens the door for integrating population genomics approaches into gene expression studies (Ekblom & Galindo 2011).

Unannotated genes

One clear advantage of RNAseq over microarrays is the identification of previously uncharacterized transcripts. RNAseq directly screens transcripts and does not rely on the design of probes from previously identified targets. However, nonmodel organisms may have limited annotation information available for species-specific genes, and there may be many unidentified genes whose functional relevance cannot be determined (Pavey *et al.* 2012). Although this problem may be mitigated by annotating sequences from homologs in closely related species using the Basic Local Alignment Search Tool (BLAST; Altschul *et al.* 1990), caution should be used when annotating genes with homologs from more divergent species, as heterologous homologs may no longer possess the same function. A long-term solution to the problem of unannotated genes is the development of databases and repositories to collect ecological annotations, which can alleviate the need for annotations derived from distantly related model organisms.

By revealing context-specific gene expression variation, novel transcripts or novel function of known transcripts, transcriptome experiments in wild settings may be the only way to infer the function of many genes present in an organism (Colbourne *et al.* 2011; Pavey *et al.* 2012). This association of gene expression with natural environmental conditions provides an 'ecological annotation' that promises to build upon existing biological process, cellular components and molecular function annotations and could be the best option for annotating genes that govern important traits (Landry & Aubin-Horth 2007; sensu the 'stress annotation' in Richards *et al.* 2012).

To assess the impact of transcriptomics in wild systems over the past decade, we summarize the primary questions that have been addressed with transcriptomic approaches in ecology. In particular, we concentrate on studies using the two main whole-genome quantification techniques: DNA microarray and RNAseq. We briefly review the literature, evaluate the experimental evidence and identify some promising questions for future research. As the availability of next-generation sequencing technology increases, ecological transcriptomics is increasingly performed using RNAseq. However, DNA microarrays represent the dominant method of transcriptome quantification over the past decade and continue to offer robust data that may still be appropriate in some systems. Therefore, clarification of the differences and appropriate applications of each type of technology is needed before we can explore the use of transcriptomics in ecology.

Microarray and RNAseq technology

Over the past 10 years, transcriptomic workflows have become increasingly refined. Microarray experiments have been the subject of multiple reviews (Allison *et al.* 2006; Crawford & Oleksiak 2007; Kammenga *et al.* 2007), and there is currently consensus on most aspects of experimental design. RNAseq is maturing, and the application to ecological studies has been discussed, but there is no current consensus on cDNA library preparation methods and data processing (Ekblom & Galindo 2011; Vijay *et al.* 2013; Wolf 2013). Previous authors explore the main differences between microarrays and RNAseq in expense, statistical analysis, bias in signals and the specific problems of using heterologous arrays and RNA pooling, which we summarize in Box 1.

DNA microarrays have been a convenient and popular tool for use in ecology, particularly because of the ease of analysis (Allison *et al.* 2006; Richards *et al.* 2009; see Box 1: *Statistical analysis*). Briefly, a DNA microarray consists of thousands of probes, representing sections of DNA code to be quantified, that are affixed to a surface. Level of expression for each gene is estimated from single probes or averages of multiple probes that are designed to target segment(s) of a gene (Ayroles & Gibson 2006). Complementary DNA (cDNA) is reverse transcribed from mRNA transcripts that are extracted from experimental material. The cDNA is labelled with a fluorescent dye before being washed over probes designed to hybridize with a specific DNA sequence. The array is scanned with lasers; relative light intensity of the fluorescent dye at a probe is proportional to the number of cDNA transcripts that are hybridizing at that particular probe (Allison *et al.* 2006; although see Box 1 *Bias in signals*). An image of the illuminated probes serves as the raw microarray data, which is preprocessed, normalized and analysed. The raw intensity data are typically first converted to a logarithmic scale (base 2) because the distribution is highly skewed: most transcripts show low expression and a minority have high expression (Ayroles & Gibson 2006).

RNAseq is a newer, increasingly popular technique for genomewide ecological transcriptomics. RNAseq uses next-generation sequencing methods to characterize RNA transcripts using high-throughput sequencing of a cDNA library to generate hundreds of thousands of fragments of DNA sequences. In the RNAseq study design phase, a user must select a next-generation sequencing platform: each platform differs in read length, sequencing depth and quality, and the impact of highly differentially expressed genes on the detection of other differentially expressed genes (Wolf *et al.* 2010; Wolf 2013; see Box 1: *Bias in signal*). Platform characteristics, coverage, costs and even available platforms, are rapidly and continually evolving, and users will need to obtain current, up-to-date information from manufacturers. Initial raw data processing requires considerable time, computing power and bioinformatics expertise, and importantly for nonmodel systems the requirements increase when assembling transcripts *de novo* (Wolf 2013). Raw RNAseq data are parsed with scripting languages due to the size of the resulting files (Malone & Oliver 2011) and then aligned to a reference genome or transcriptome via software such as the Tuxedo Suite Tools (Trapnell *et al.* 2012), or assembled *de novo* via software such as Trinity (Grabherr *et al.* 2011).

Studies reviewed: what have we learned?

Over the past decade, both RNAseq and microarrays have made critical contributions to ecology. We used a systematic review approach (Doerr *et al.* 2014) to characterize the development of this field in an objective, repeatable fashion. We found 575 studies published

between 2004 and 2013 through the Web of Science database (see Appendix S1, Supporting information). Each study shared the wild card search term 'transcriptom*' which initially returned 307 000 studies. We further refined the search by choosing studies only from the Web of Science research areas 'evolutionary biology' and 'environmental science ecology' which returned 2303 empirical studies. From these studies, we concentrated on ecological and evolutionary studies and manually excluded studies that primarily referred to toxicology, agriculture or other applications in environmental science without an obvious ecological context. We also excluded meta-analyses and studies that performed analyses on previously generated data. Rather than classifying studies by the ecological phenomenon or study organism, we identified a more general summation based on three questions that reflects how ecological transcriptomics has made a transition over the past 10 years from largely descriptive investigations to those that are more functional and mechanistic (Fig. 1): (i) How much variation in gene expression is there in natural populations and how is it structured? (ii) How do environmental stimuli affect gene expression? (iii) How does variation in gene expression translate into phenotype? In the following, we elaborate on how each of these questions has been addressed. Note that some studies addressed more than one of these questions and were classified into multiple categories (Fig. 2; Appendix S1, Supporting information). We also identified which technology (55% of the total were DNA microarray, 45% were RNAseq) each study used, and whether organisms in each study were reared in or acclimated to laboratory conditions before sampling, or were sampled from a field setting.

How much variation is there in gene expression and how is it structured?

One of the critical questions for understanding the importance of any trait in ecology and evolution is: How much variation exists in natural populations and how is it structured? Gene expression is highly variable, and transcription levels vary within multiple biological scales: within and among individuals, populations and species (Whitehead & Crawford 2006a,b; Crawford & Oleksiak 2007). Within individuals, gene expression varies between tissues and even among cell types within the same tissue (Birnbaum *et al.* 2003; Whitehead & Crawford 2006b). Gene expression further varies temporally, fluctuating with developmental time, day–night cycles and life history events (Aubin-Horth & Renn 2009; Francesconi & Lehner 2013). A majority of studies that we reviewed (66%) quantified variation in gene expression in one or more natural populations, even if this was not the main focus of the study. Studies developing transcriptome resources for the first time appear in this category, as they often represent a 'first look' at a nonmodel organism's transcriptome. This category is the most descriptive in nature, and several of the studies in this category represent some of the earliest and most fundamental questions in ecological transcriptomics: How does gene expression vary within individuals, within and among populations, and among species? Understanding existing levels of variation in gene expression is an important question because changes in gene expression may allow organisms to respond to novel stressors and variation in gene expression may translate into phenotypic variation (Oleksiak *et al.* 2002; Whitehead & Crawford 2006a; Whitehead 2012).

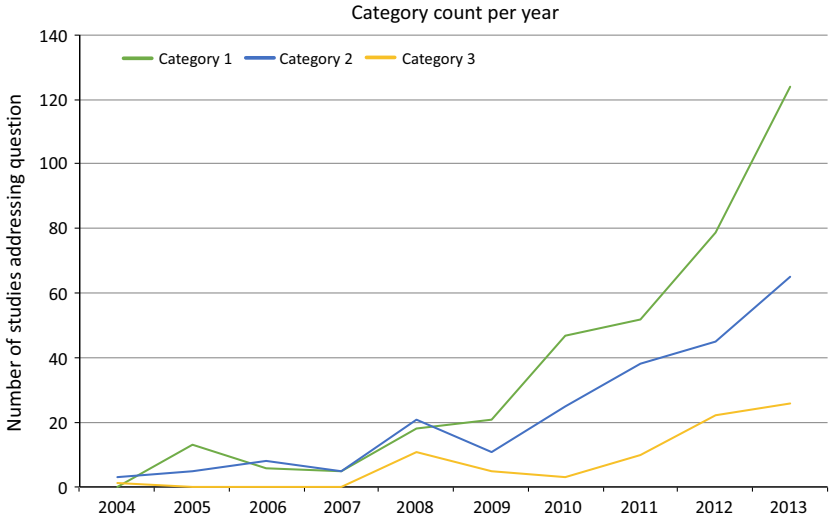

Fig. 1 Number of studies each year addressing each of the question categories. Question categories: 1 corresponds to studies that investigated the extent and structure of natural variation, 2 corresponds to studies that investigated organismal response to stimulus and 3 corresponds to studies that investigated the effect of differential gene expression on phenotype.

Variation in gene expression can be shaped by evolutionary processes

Variation in gene expression is potentially heritable and may be acted upon by natural selection. As with any other trait, the structure of gene expression variation within and among natural populations may reflect both adaptive and nonadaptive processes (Oleksiak *et al.* 2002; Whitehead 2012). Expression variation may be facilitated by regulatory elements or epigenetic mechanisms that alter gene expression even before genetic variants arise in the population (West-Eberhard 2003). Therefore, population-level differences in expression may reflect the early processes that underlie adaptive divergence (e.g. in Oleksiak *et al.* 2002; Derome *et al.* 2006; Jeukens *et al.* 2010). To quantify the expression variation that is correlated with the early stages of divergence within populations, Derome *et al.* (2006) used microarrays to investigate transcriptional differences between differentiated 'normal' and 'dwarf' types of the fish *C. clupeaformis*. Previous work demonstrated that the 'normal' and 'dwarf' types had physiological variation in swimming activity which might be partly due to expression of genes related to energy metabolism (Bernatchez & Dodson 1985). The authors sampled individuals of both types from two populations and compared transcription in muscle tissue. They found 51 differentially expressed genes between the two types, which as hypothesized, were primarily annotated as being involved in energetic metabolism. The authors were able to detect expression variation that may reflect adaptive divergence between two sympatric subpopulations of *C. clupeaformis* and identify candidate genes for future analyses.

To specifically examine how adaptive or neutral processes have contributed to variation in gene expression in natural populations, several studies have adopted classic approaches like McDonald-Kreitman tests, Q_{st}–F_{st} tests and quantitative trait loci (QTL) mapping. In particular, the salmonid fishes have been a useful system to tease apart the importance of selection and drift on gene expression in natural populations. In one of the first applications of Q_{st}–F_{st} analysis to transcriptomics, Roberge *et al.* (2007) used both a genome scan and Q_{st}–F_{st} analysis to identify genes whose transcriptional profiles (assessed via microarray) had been shaped by selection in two diverging subpopulations of salmon (*Salmo salar*). Q_{st} usually quantifies the amount of variation in quantitative traits in populations (Spitze 1993), which can be compared to variation at neutral loci (F_{st}) to identify selection and drift in phenotypic divergence (Koskinen *et al.* 2002; Roberge *et al.* 2007). The authors adapted the Q_{st} framework to gene expression by treating gene expression as a quantitative trait and estimated transcription level Q_{st} values for 1044 genes with transcriptional profiles that were highly heritable. After narrowing their search to genes that were divergent between the two populations and testing for a neutral model of genetic variation, the authors identified only 16 genes that probably diverged between the subpopulations due to selection, rather than neutral processes. Conversely, they found 11 divergent genes whose expression did not reflect directional selection, but may be influenced by neutral processes, suggesting a role for drift in shaping the expression patterns of *S. salar* (Roberge *et al.* 2007). This combination of natural populations, captive breeding and genomic techniques is a powerful model for investigators attempting to disentangle the role of selection and neutral processes in natural populations.

Fig. 2 Venn diagram of question categories in studies examined. Black numbers represent question categories. Red numbers indicate the number of studies in the question category. Overlapping areas represent studies that were represented by more than one question category. See Appendix S1 (Supporting information) for details of the 575 studies summarized here.

An expression quantitative trail loci (eQTL) mapping approach is another useful tool for identifying important expression variation within populations (Brem *et al.* 2002; Wittkopp 2007). eQTL analyses treat gene expression as a quantitative trait and use classic QTL methods to map genetic loci that underlie variation in gene expression. Further, eQTL mapping can identify whether genes are modified by *cis*- or *trans*-regulation (Brem *et al.* 2002; Wittkopp 2007; Derome *et al.* 2008; Whiteley *et al.* 2008). Although eQTL is usually performed in model organisms, Derome *et al.* (2008) were able to make use of

a linkage map generated for *C. clupeaformis* to identify 34 transcripts that may play a role in the ongoing divergence between the two types, and thus may be under selection (Derome *et al.* 2008). The power of eQTL and Q_{st}–F_{st} analyses can be leveraged in natural environments to identify the effects of evolutionary forces on transcripts of interest that would otherwise be unidentifiable under laboratory conditions.

Macroevolutionary patterns of variation in gene expression: the comparative method

Because neutral and adaptive processes contribute to variation in gene expression, examining the contributions of these processes to differential expression patterns in a phylogenetic comparative framework may contribute to our understanding of divergence at the species or higher taxonomic levels (Wittkopp 2007; Whitehead 2012; Schraiber *et al.* 2013). Comparisons between closely related species can identify genes (e.g. Lai *et al.* 2006; Chelaifa *et al.* 2010), or changes in regulatory pathways (e.g. Brem *et al.* 2002; Schraiber *et al.* 2013) that may have been important in speciation events and test hypotheses about the importance of response to environmental challenges to speciation (Wittkopp 2007). For example, Chelaifa *et al.* (2010) used microarrays to explore the transcriptomes of *Spartina maritima* and *Spartina alterniflora*, two recently diverged sister species that occupy overlapping niches. The authors found differential expression in 13% of the transcriptome, including genes annotated as transporter genes, developmental genes and cellular growth genes. These divergent expression patterns may play a role in allowing the species to inhabit their different ecological niches and in shaping differentiated phenotypes (Chelaifa *et al.* 2010).

Although this category is the most descriptive of the three categories examined, studies in this category have identified gene expression variation across time, space and phylogenetic distance in complex natural environments. Additionally, advanced approaches, such as Q_{st}–F_{st}, have been able to quantify the influence of selection, drift or bottleneck events on the evolution of gene expression. Further application of these approaches to new or existing data sets may shed light on the relative influence of these factors in diverging or recently diverged populations. Additionally, the use of captive breeding populations may aid in the development of resources, such as genetic maps (as in Derome *et al.* 2008). These resources can then be combined with sampling in natural systems for more powerful discovery of the forces that shape diversity in gene expression.

How do environmental stimuli affect gene expression?

A central theme in molecular ecology explores the mechanisms by which organisms respond to their environment, a question which has taken on increasing importance in the shadow of global climate change (Nicotra *et al.* 2010; Meier *et al.* 2014; Palumbi *et al.* 2014). This question is the logical next step in ecological studies after describing patterns of variation, and just under half of the studies we reviewed (41%) addressed how environmental stimuli affect gene expression. Our definition of environmental stimulus includes, but is not limited to, abiotic stress (such as temperature or pollution; e.g. Chapman *et al.* 2011; Whitehead *et al.* 2012; Palumbi *et al.* 2014), environmental heterogeneity in time or space (e.g. Richards *et al.* 2012; Meier *et al.* 2014), host–parasite interactions (e.g. Webster *et al.* 2011), and other, potentially selective biotic and abiotic interactions.

Natural environmental fluctuations impact gene expression

Ten years of transcriptomics in the wild have described the significant impacts that stress response can have on many categories of genes, but have also shown that transcription may be affected by even small changes in environment (Krasnov *et al.* 2005; Richards *et al.* 2012). For example, Richards *et al.* (2012) used DNA microarrays to explore how gene expression changes through development and in response to environmental conditions. The authors found that variation in gene expression in two accessions of *A. thaliana* grown in field conditions was equally explained by differences in accession and developmental status and that temperature and precipitation significantly predicted expression. The authors also created a 'stress annotation' of the *A. thaliana* genome based on published microarray studies to show that genes previously annotated as stress-related were expressed during the life cycle of the organisms even under normal field conditions (Richards *et al.* 2012). Using a simple design and several environmental measurements, this study teased apart the influence of development and abiotic environmental variables in a complex natural setting. In addition to identifying the molecular level basis of response to environmental challenges, this type of study elucidates genotype-by-environment interactions at the molecular level (Wray 2013) and reveals the molecular mechanisms of phenotypic plasticity (e.g. Chapman *et al.* 2011; Gunter *et al.* 2013; Whitehead *et al.* 2013).

Gene expression responds to extreme events or environmental stress

Transcriptomic data can also provide insight into mechanisms of organismal response to specific pollution events, stresses or climatic conditions in wild populations. For example, Whitehead *et al.* (2013) recently used a combination of RNAseq and microarrays to examine the molecular and physiological response to osmotic challenge in two species of killifish, *F. heteroclitus* and *Fundulus majalis*, which vary in tolerance to changes in salinity. They reported rapid divergence in gene expression between the two species in response to osmotic stress. The authors also found a greater capacity for the morphological remodelling of gills in *F. heteroclitus*, and suggested that the expression and correlated morphological variation they observed could have played a role in the divergence of *F. heteroclitus* from *F. majalis* allowing it to inhabit a broader range of osmotic environments (Whitehead *et al.* 2013). In another study, Chapman *et al.* (2011) found that environmental conditions (pH and temperature) were the primary drivers of differential transcription in populations of the eastern oyster (*Crassostrea virginica*), as opposed to the minimal effects attributed to pollutants (metals and organic contaminants). The authors were able to develop statistical models of predictive value that were parameterized by data from natural populations, rather than simulated data or data obtained from controlled studies.

Practical implications of rapid response of gene expression to environment

The responsiveness of gene expression has the practical implication that sampling requires consistent handling with minimal exposure to unwanted stimuli before and during collection. When possible, samples should be collected at approximately the same time and flash frozen immediately upon collection. RNA preservation additives, such as RNAlater (Qiagen), may also prevent RNA degradation between sampling times. To minimize batch effects, each sample must be handled consistently to prevent differential environmental stimuli from having effects on the samples. As an alternative to immediate freezing of samples, some experiments have attempted to minimize intra-individual variations by letting organisms sit in a uniform space for a period of time before samples are sacrificed and frozen. This may be necessary when logistics prevent immediate preservation, but may introduce unwanted variation into the sample populations by introducing additional handling, feeding and other environmental variables that may affect gene expression.

Another technical difficulty is that transcriptome assays represent only a 'snapshot' of gene expression at a particular moment: temporal variation is necessarily controlled for by careful sampling during the same time of day, climate cycle or time during a stimulus response to get relevant comparisons between groups. In all of the studies addressing environmental response, ecological experimental design and analysis allowed investigators to discriminate between the effects of multiple, complex environmental inputs in wild populations. However, most studies used data obtained from a single time point to describe organismal response to stimulus. Unfortunately, this methodology condenses the temporo-spatial variation in a transcriptional response into a single time point. Future transcriptome studies should explore temporal levels of variation, as temporal differences in gene expression may help to pinpoint the primary regulatory loci that allow organisms to modulate gene networks and subsequently phenotype in response to endogenous or exogenous stimuli (West-Eberhard 2003; Aubin-Horth & Renn 2009; Dalziel *et al.* 2009; Johnson *et al.* 2012). Gene expression response may involve variation in regulatory loci early in the response, followed by a generalized downstream response in other genes (Wittkopp 2007; de Nadal *et al.* 2011). Alternatively, gene networks may respond sequentially to a complex environmental stimulus as various conditions are met (Aubin-Horth & Renn 2009).

A more nuanced understanding of expression: the use of time-course studies

Aubin-Horth & Renn (2009) suggest using a time-course approach to understand temporal patterns in transcription. Time-course designs allow for a better description of the entirety of a transcriptional response and assist in ecological annotation by teasing apart general response genes from major regulatory genes. The replication normally associated with ecological studies, combined with the replication required for surveying multiple time points, may quickly become overwhelming in both cost and scale. However, measuring behaviour of previously identified candidate genes with qPCR (rather than using whole-genome assays) may provide useful data on temporal response without incurring prohibitive costs. A recent study used this approach to explore the role of candidate genes in the environmentally responsive network that underlies the diet-induced plasticity of the lower jaw in cichlids (Schneider *et al.* 2014). The authors identified a pattern across development in response to diet: first mechanically responsive genes, then osteoblast differentiation genes, then matrix-related genes were differentially expressed. In addition, the time-course design allowed

for identifying the so-called immediate early genes (e.g. AP1) that influenced expression at different levels of the regulatory cascade.

Systems biology methods isolate the importance of specific environmental factors

As ecological transcriptomics identifies the stimuli that affect patterns of transcription, the effects of climatic and meteorological fluctuations on loci of interest should become an important focus of study. Climatic variation can have large impacts on gene expression (e.g. Richards *et al.* 2012) and is a critical component in surveying organismal response to climate change. When response to these variables is not the primary interest, the incorporation of climate data will allow investigators to control for the impact of climate across taxa and habitats. When these variables are of interest, a clear understanding of climate variables will allow researchers to make predictions about which factors (temperature, rainfall, CO_2) are driving variation in gene expression. Nagano *et al.* (2012) provide an example in rice (*Oryza sativa*) where the authors used microarrays to model the expression of each gene in response to climate variables at time points across development. In soya bean (*Glycine max*), Leakey *et al.* (2009) used microarray data and physiological measurements to determine the mechanism of respiratory regulation in a free-air CO_2 enrichment (FACE) experiment. The authors found variation in metabolic gene expression during periods of elevated CO_2 and a concurrent increase in night-time respiration.

Although these experiments used agricultural species, their approaches are useful to determine the impact of climate on wild populations. Aikawa *et al.* (2010) used a similar design to model the expression of a single gene in the flowering time network over 2 years in *Arabidopsis halleri*. The authors found that at any time point, expression follows cues from the prior 6 weeks of temperature data. Higher-throughput expression analysis may reveal other patterns of response to specific environmental factors within the flowering time network. These studies often use systems biology approaches of unsupervised classifications such as principal components analysis and clustering (e.g. Richards *et al.* 2012), and regression analysis for each gene (e.g. Nagano *et al.* 2012), to identify specific environmental factors that impact gene expression in an ecological context (Richards *et al.* 2009; Shimizu *et al.* 2011).

How does gene expression affect phenotype?

For gene expression to play a functional role in ecology, it must affect phenotype. Characterizing the relationship between gene expression and phenotype provides critical insight into understanding how ecological and evolutionary processes, such as adaptive divergence, take place at the molecular level and exert influence on phenotype. Despite this imperative, a minority of the studies we reviewed made the connection between gene expression and phenotype (15%); this was the most underrepresented research question, which has increased only in recent years (Fig. 1). The relationship between gene expression and phenotype is complex, as gene expression may have interactive effects with other larger scale systems, such as the proteome and metabolome and may not immediately impact fitness (Crawford & Oleksiak 2007; Dalziel *et al.* 2009; Rees *et al.* 2011). Most studies in this category relied on patterns of correlation between expression of functional genetic elements and the production of a particular phenotype (including proteins, metabolites or traits). In one example, Aubin-Horth *et al.* (2012) linked differential expression in genes that are involved in stress response to behavioural variation in *Gasterosteus aculeatus*. The authors were able to demonstrate a correlation between expression levels in candidate genes with variation in boldness and aggressiveness (Aubin-Horth *et al.* 2012). Few studies have confirmed the causal relationship between functional elements and phenotype through additional protein- or metabolism-based assays (Rees *et al.* 2011; Whitehead *et al.* 2011), by knocking out genes of interest (Dowen *et al.* 2012; Richards *et al.* 2012), or through transgenic expression of genes of interest (Kobayashi *et al.* 2013).

Alternate phenotypes reveal expression differences that translate into phenotypic variation

In organisms with alternate phenotypes, analysis of differential gene expression can help to explain the processes by which the phenotypes diverge and resources are allocated to create the alternative types (Derome *et al.* 2008; Cardoen *et al.* 2011; Gunter *et al.* 2013; Schneider *et al.* 2014). For instance, honeybee workers appear in two phenotypes: nonaltruistic reproductive forms and altruistic, nonreproductive forms. Cardoen *et al.* (2011) hypothesized that environmental signals, received by the worker bees, control the activation of the ovaries. The authors found 1292 genes, involved in multiple metabolic pathways, which were differentially transcribed between the two phenotypes and identified candidate genes which were potentially linked to the phenotypic differentiation between nonaltruistic reproductive forms and altruistic reproductive forms. In another study, Filteau *et al.* (2013) used a weighted gene co-expression network analysis to identify the architecture of gene networks that were divergent between the

two previously discussed 'normal' and 'dwarf' types of the fish *C. clupeaformis*. The authors sampled brain and muscles tissue and found 14 and 17 co-expression modules, respectively, that differed between the two types. A gene network-based approach may lend additional functional information about morphological differentiation and divergence when gene annotation information is insufficient, and provides further, stronger correlation between gene expression and phenotype.

Other studies have used large-scale environmental disturbance as a natural experiment to reveal genes that contribute to phenotypes such as pollution or drought tolerance. Natural disturbance events encapsulate numerous biotic and abiotic interactions that may be difficult or impossible to model under controlled conditions. For example, Whitehead *et al.* (2011) used microarrays to examine the impact of the *Deepwater Horizon* oil spill on the transcription and physiology of the Gulf killifish (*Fundulus grandis*). The authors sampled individuals from six different field sites across three time points and identified more than 1500 genes that were differentially expressed in response to oil stress. Additionally, individuals exposed to oil showed altered gill morphology. One responsive gene, *cyp1a*, is known to cause developmental abnormalities and decrease larval survival. A follow-up study under controlled conditions confirmed that the CYP1A protein was expressed in response to oil exposure, and this expression was particularly localized to areas of the gills, which showed altered morphology. This combined approach provided a quantitative link between differential transcription, differential protein expression and individual phenotype. Leveraging the natural 'treatment and control' design created by the oil spill allowed the authors to test hypotheses about organismal response in situ rather than relying on extrapolations from laboratory studies.

A few studies have combined transcriptomics with controlled studies of evolution to explore the molecular mechanisms of adaptation over ecological and evolutionary time. For example, Dhar *et al.* (2011) monitored adaptation to salt stress in *Saccharomyces cerevisiae* using both microarrays to examine expression differences and DNA sequencing to quantify sequence changes. Adaptation was quantified by measuring changes in population growth rate, a measure of fitness. Adaptive changes were correlated with a single SNP and differences in genome size, both of which may have contributed to the differential expression of 1431 genes. This indicated that the evolution of gene expression may have played an important role in adaptation to this stress (Dhar *et al.* 2011). Controlled studies of evolution allow investigators to track the effects of expression changes and phenotypic differentiation over evolutionary time, and quantify outcomes such as population divergence and adaptation.

Moving from correlation to causation

Ultimately, it will be imperative to move beyond correlating patterns of gene expression variation with patterns of trait variation. Incorporating manipulations of transcription (through e.g. transgenics, RNAi or CRISPER/Cas) into transcriptome studies addresses the relative lack of data on whether differential transcription can 'trickle up' to affect phenotype and ultimately populations (Ungerer *et al.* 2007; Dalziel *et al.* 2009; Pavey *et al.* 2012). Identifying the impacts of controlled changes in transcription in concert with ecologically relevant traits in a natural setting will refine our understanding of well-known genetic pathways of interest in model and nonmodel systems and reveal how these pathways may have diversified across taxa. For example, based on annotation of flowering time genes in the model plant *A. thaliana*, Kobayashi *et al.* (2013) identified transcriptional changes in homologs of a floral pathway integrator (SbFT) and a floral repressor (SbSVP) before a community level mass flowering event in the tropical tree *Shorea beccariana*. The function of these genes was confirmed using transgenic *A. thaliana*: when compared with the wild type, the transgenic *A. thaliana* overexpressing SbFT showed early flowering, whereas late flowering was observed for those overexpressing SbSVP. Another study by Zhu *et al.* (2008) examined gene expression data from segregating populations of yeast to construct gene networks. The authors used co-expression data, along with transcription factor-binding site and protein–protein interaction information, to build gene networks. These gene networks were then linked to phenotype via eQTL analysis and confirmed through analysis of gene knockout lines, allowing the authors to describe the causal effects of expression networks on phenotype. Although these two studies used transgenic individuals and knockout lines to confirm gene function, other manipulations such as RNAi or CRISPER/Cas have been used for emerging nontraditional model species (e.g. Hwang *et al.* 2013) and may be easily applied to nonmodel species of ecological interest. By focusing on this type of enquiry, ecological transcriptomics can continue its progression from a discipline that describes pattern, to one that elucidates process and informs ecological and evolutionary theory.

Transcriptomics in the future: where do we go from here?

Ten years of ecological transcriptomics have yielded descriptions of transcriptional variation in natural pop-

ulations of a variety of organisms and in response to a variety of stimuli. We have described some of the reasons that microarrays or RNAseq have been appropriate based on study system and research question (Box 1). As the field progresses, future studies, particularly of organisms with no genomic resources, will most likely rely on RNAseq (already 45% of studies reviewed use RNAseq), but microarrays may still offer valuable data, depending on the study system, the question and design issues.

A unifying workflow

Although microarrays represented the primary method of whole-genome transcription quantification for the past decade, RNAseq studies have increased greatly in recent years, and previous statistical and technical limitations are rapidly being addressed. Given the advantages of each of the major transcriptomic technologies (see Box 1), microarrays and RNA-seq may be combined to test hypotheses about the importance of global gene expression patterns in natural populations (Malone & Oliver 2011). If a commercial or custom microarray is already designed for a given species, it may be easily applied to a new study on the same species, unless the question explicitly involves differences in expression of candidate genes that were not included in the array design. However, the user must decide whether the available probes are relevant for the question being investigated. For instance, a microarray based on locust (*Schistocerca gregaria*) ganglia (Badisco *et al.* 2011) will not be able to identify all relevant transcripts in other locust tissue types. A nonexhaustive list of commercially available and custom microarrays (Table 1) gives an indication of the wide range of taxa represented by existing microarrays. It is important to remember that unlike RNAseq, microarrays cannot give information about previously unidentified transcripts, transcript sequence or alternate isoforms. However, for quantifying variation in response to stimuli or surveying patterns of gene expression in the wild, microarrays are still a useful and viable choice, especially if they are already available.

In organisms without previous genomic resources, RNAseq will most efficiently quantify standing transcriptomic variation in a species and identify gene targets of interest. Once genes and gene networks of interest have been identified in the study organism, these transcripts could be used to generate annotations, provide information about alternative gene isoforms or to construct a specialized microarray for future studies. In an example of this integrated technique, Vera *et al.* (2008) used high-throughput RNAseq and de novo assembly of the Glanville fritillary butterfly (*Melitaea cinxia*) transcriptome from 80 individuals across eight populations. The sequencing results were used to construct a microarray, which was used in two follow-up studies, one examining differential gene expression between older and more recently established populations of *M. cinxia* (Wheat *et al.* 2011) and another investigating heritable gene expression variation in *M. cinxia* larval development (Kvist *et al.* 2013). This series of investigations leveraged the ability of RNAseq to characterize previously unexplored genomes to create a robust microarray for follow-up experiments. This is a powerful experimental pipeline for ecological transcriptomics of nonmodel organisms when genes of interest have already been identified with RNAseq. If the identification of novel transcripts is still of interest, or when generating ecological annotations, RNAseq can identify previously unknown transcripts while still providing data on expression variation. This can be a bioinformatics challenge and many researchers will benefit from commercial options for bioinformatics.

While it is now possible for transcriptomics to probe genomewide patterns, many groups still use single- or multilocus assays, in the form of single- and multilocus qPCR, that quantify the expression of a select subset of genes used to probe the transcriptome. These techniques are important in surveying ecologically relevant candidate genes of interest for disease and response to environmental conditions (e.g. Aikawa *et al.* 2010; Schneider *et al.* 2014) and are important in the validation of gene expression. The relatively low cost and high reliability of qPCR makes it valuable for validating genomewide expression techniques for two purposes. First, technical validation confirms that the platform used to survey genomewide expression is accurate. This validation is commonly performed after a genomewide survey of expression. A second, less common biological validation confirms that the phenomenon of interest actually causes the observed variation in gene expression or vice versa (Kammenga *et al.* 2007). Ideally, biological validation of gene function uses independent biological samples to confirm the up- or downregulation of genes in response to a given treatment or condition of interest. Therefore, although we did not include studies that relied solely on qPCR in our survey of transcriptomics, the use of qPCR for confirmation of the expression of genes of interest is essential.

Future directions for enquiry

Thanks to the power of genomewide expression studies, we are now at a point in our understanding of genome function where we can and should move beyond telling single-gene stories and start assembling a systems-level

Table 1 DNA microarrays provided commercially by Agilent and Affymetrix, and custom built

Agilent	Affymetrix	Custom	
Arabidopsis thaliana	*A. thaliana*	*Acropora millepora*	*Locusta migratoria*
Bovinae spp. (bovine)	*Bovinae* spp. (bovine)	*Acropora palmata*	*Loxodonta africana*
Brassica spp.	*Caenorhabditis elegans*	*Ambrosia artemisiifolia*	*Megachile rotundata*
C. elegans	*Callitrichidae* spp. (Marmoset)	*Amphiura filiformis*	*Melitaea cinxia*
Canis spp. (canine)	*Canis* spp. (canine)	*Anemonia viridis*	*Meliteaea cinxia*
Culicidae spp. (mosquito)	*Danio rerio* (Zebrafish)	*Anopheles gambiae*	*Metarhizium robertsii*
D. rerio (Zebrafish)	*Drosophila melanogaster*	*Apis mellifera carnica*	*Montastraea faveolata*
D. melanogaster	*Equus ferus* (horse)	*Azospirillum brasilense*	*Mytilus californianus*
E. ferus (horse)	*Felidae* spp. (feline)	*Brugia malayi*	*Neotoma lepida*
Gallus gallus (chicken)	*G. gallus* (chicken)	*Calanus finmarchicus*	*Neurospora crassa*
Gossypium spp. (cotton)	*Glycine max* (soybean)	*Carpodacus mexicanus*	*Neurospora tetrasperma*
Homo sapiens	*H. sapiens*	*Coregonus cupleaformis*	*Onthophagus taurus*
Hordeum vulgare (barley)	*Macaca mulatta* (rhesus)	*Crassostrea gigas*	*Ostreococcus tauri*
Leporidae spp. (rabbit)	*Medicago* spp.	*Crassostrea virginica*	*Petrolisthes cinctipes*
M. mulatta (rhesus)	*Mus musculus*	*Daphnia magna*	*Picea glauca*
Magnaporthe spp.	*Oryza sativa* (rice)	*Daphnia pulex*	*Pimephales promelas*
Medicago spp.	*Ovis aries* (sheep)	*Dreissena polymorpha*	*Pinus pinaster*
M. musculus (mouse)	*Saccharomyces cerevisiae*	*Drosophila mojavensis*	*Pinus taeda*
Nicotiana tabacum (tobacco)	*Schizosaccharomyces pombe*	*Eucalyptus grandis*	*Platichthys flesus*
O. sativa (rice)	*Suidae* spp. (porcine)	*Folsomia candida*	*Ruditapes philippinarum*
O. aries (sheep)	*Taeniopygia guttata* (Zebra Finch)	*Fundulus heteroclitus*	*Ruditapes phillippinarum*
Rattus norvegicus (rat)		*Gadus morhua*	*Salmo salar*
S. cerevisiae		*Gasterosteus aculeatus*	*Salvelinus fontinalis*
Salmonidae spp. (salmon)		*Helianthus annus*	*Schistocerca gregaria*
Solanum lycopersicum (tomato)		*Karenia brevis*	*S. lycopersicum*
Suidae spp. (porcine)		*Lagopus lagopus*	*Solenopsis invicta*
Triticum spp. (wheat)		*Lagopus lagopus scoticus*	*Tribolium castaneum*
Xenopus spp.		*Laternula elliptica*	*Tursiops truncatus*
		Lepeophtheirus salmonis	

understanding of how organisms respond to environmental challenges (Wittkopp 2007; Richards *et al.* 2009; Schraiber *et al.* 2013). One trend that emerges from our survey of the literature is that as the molecular revolution in ecology progresses, ecological transcriptomics is moving from a largely descriptive discipline to one which identifies the causal elements of phenotypic change in wild populations (Aubin-Horth & Renn 2009; Andrew *et al.* 2013; Fig. 1). In the first 10 years of ecological transcriptomics work, authors rarely followed up to test the predicted importance of gene expression variation on response to phenotype. Instead, most relied heavily on Gene Ontology types of analyses to infer relevant biological function without experimentally confirming that these inferences were true. Validating our findings from transcriptome studies may require assays at other molecular levels (e.g. Rees *et al.* 2011; Whitehead *et al.* 2011), and the use of knockouts (e.g. Dowen *et al.* 2012; Richards *et al.* 2012) or transgenic organisms (e.g. Kobayashi *et al.* 2013) grown in ecologically relevant experiments. Taking full advantage of the power of transcriptomics in ecology requires the integration of robust experimental designs and a synthetic approach that includes molecular, morphological, physiological or behavioural measurements at other levels of biological organization (Vasemagi & Primmer 2005; Richards *et al.* 2009).

As molecular ecology shifts from describing correlation to identifying causation, ecological transcriptomics will help elucidate the role of genomic elements that precede, regulate and follow transcriptional modulation. Understanding the role of different genomic elements will allow investigators to more fully examine the pathways through which differential gene expression modulates phenotypic traits. Epigenetic mechanisms such as DNA methylation, which can result in mitotically or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence, have been correlated with a variety of ecological stimuli across taxa and have potentially heritable effects on phenotype (Kilvitis *et al.* 2014). As a proximate cause of transcriptional variation, DNA methylation assays may add functionally relevant information to discriminate between environmental stimuli. For example, Dowen *et al.* (2012) identified changes in the model plant *A. thaliana* in response to bacterial pathogen, avirulent

bacteria and the defence hormone salicylic acid under laboratory conditions. The authors identified differentially methylated regions throughout the genome and used RNAseq to identify differentially transcribed genes located near differentially methylated regions. The authors were thus able to identify stress-response genes whose transcription was altered by differential methylation in response to ecologically relevant stimuli. Studies have shown that DNA sequence-based differences can be context dependent, but DNA methylation is even more labile to environmental influence, and natural settings may induce alternative methylation profiles that would not be visible in controlled settings. Combining genomewide expression surveys of wild populations with DNA methylation quantification may allow a more complete picture of the genetic architecture of environmental response.

Proteomics may also provide an avenue for linking transcriptional variation to larger biological processes (Vasemagi & Primmer 2005; Diz *et al.* 2012). Quantifying protein expression may add functional information about a gene's response as understanding actual translation to protein product is crucial to mapping the ultimate effects of differential gene expression on phenotype. Proteins may also be modified post-translation to enhance or temper the cellular impact of differentially expressed genes, making proteomics an important tool for measuring the final impact of gene regulation on phenotype (Diz *et al.* 2012). Rees *et al.* (2011) examined gene and protein expression in *Fundulus* species to correlate response at these two levels among three populations of *Fundulus*. The authors found that although mRNA is generally positively correlated with protein expression, the relationship is nonlinear. Regulatory mechanisms among proteins may alter or enhance gene expression differences (Rees *et al.* 2011). Further studies integrating these additional molecular markers, along with functional and phenotypic analyses, may allow for a better mechanistic explanation of heritable differences between populations.

As the field of ecological transcriptomics now represents a major data stream in molecular ecology, investigators and institutions must build an infrastructure to support increased gene annotation. Traditional model organisms, such as *Mus musculus*, *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *A. thaliana* have a wide array of genomic resources available to them, including gene annotations, which identifies the function of a putative transcript, and predicted gene interactions. However, ecologists are typically interested in nonmodel organisms, which almost always lack these genomic resources. Despite increased access to genomic tools over the past decade, molecular ecologists are still limited in their ability to use genomic data because of a lack of information on relevant genes (Pavey *et al.* 2012; Andrew *et al.* 2013). Although there has been a growing push towards an ecological annotation of genes (Landry & Aubin-Horth 2007; Aubin-Horth & Renn 2009; Pavey *et al.* 2012; Andrew *et al.* 2013), investigators working on nonmodel organisms are still largely limited to using annotations from their closest model relative. As genetic distance increases, the likelihood increases that a putative ortholog, or a gene that is related by vertical descent, has diverged and an annotation from a model organism is not accurate. Further, even in model organisms, gene annotations are not available for the entire genome. Ecological gene annotations from wild species may greatly enhance annotations from related model organisms. As mentioned earlier, genes that are species- and context-specific may be vitally important in explaining ecological processes and interactions (Colbourne *et al.* 2011). A long-term solution to alleviate the problem of poorly annotated genes is the creation of a database for proposed ecological annotations (Pavey *et al.* 2012). While we have made much progress in ecology by examining traits and behaviours of individuals within and among populations, the integration of molecular techniques into ecology allows investigators unprecedented ability to examine the mechanistic underpinnings of the diverse phenotypes that contribute to phenotypic variation and rapid response to environment. An enhanced focus on ecological transcriptomics promises to contribute a powerful component to our understanding of the molecular basis of ecological interactions and evolutionary processes.

Acknowledgements

We thank Christy Foust, Holly Kilvitis, Marta Robertson, Larissa Williams and Koen Verhoeven and three thoughtful anonymous reviewers for valuable feedback on the manuscript. This work was supported by the University of South Florida (MA & CLR).

References

Aikawa S, Kobayashi MJ, Satake A, Shimizu KK, Kudoh H (2010) Robust control of the seasonal expression of the Arabidopsis FLC gene in a fluctuating environment. *Proceedings of the National Academy of Sciences of the USA*, **107**, 11632–11637.

Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J (2014) Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid Arabidopsis. *Nucleic Acids Research*, **42**, e46.

Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics*, **7**, 55–65.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Andrew RL, Bernatchez L, Bonin A *et al.* (2013) A road map for molecular ecology. *Molecular Ecology*, **22**, 2605–2626.

Aubin-Horth N, Renn SC (2009) Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Molecular Ecology*, **18**, 3763–3780.

Aubin-Horth N, Deschênes M, Cloutier S (2012) Natural variation in the molecular stress network correlates with a behavioural syndrome. *Hormones and Behavior*, **61**, 140–146.

Ayroles JF, Gibson G (2006) Analysis of variance of microarray data. *Methods in Enzymology*, **411**, 214–233.

Badisco L, Huybrechts J, Simonet G *et al.* (2011) Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database. *PLoS ONE*, **6**, e17274.

Bernatchez L, Dodson JJ (1985) Influence of temperature and current speed on the swimming capacity of lake whitefish (*Coregonus clupeaformis*) and cisco (*C. artedii*). *Canadian Journal of Fisheries and Aquatic Sciences*, **42**, 1522–1529.

Birnbaum K, Shasha DE, Wang JY *et al.* (2003) A gene expression map of the Arabidopsis root. *Science*, **302**, 1956–1960.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.

Buckley BA (2007) Comparative environmental genomics in non-model species: using heterologous hybridization to DNA-based microarrays. *Journal of Experimental Biology*, **210**, 1602–1606.

Cardoen D, Wenseleers T, Ernst UR *et al.* (2011) Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Molecular Ecology*, **20**, 4070–4084.

Chapman RW, Mancia A, Beal M *et al.* (2011) The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Molecular Ecology*, **20**, 1431–1449.

Chelaifa H, Mahe F, Ainouche ML (2010) Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Molecular Ecology*, **19**, 2050–2063.

Colbourne JK, Pfrender ME, Gilbert D *et al.* (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**, 555–561.

Crawford DL, Oleksiak MF (2007) The biological importance of measuring individual variation. *Journal of Experimental Biology*, **210**, 1613–1621.

Dalziel AC, Rogers SM, Schulte PM (2009) Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Molecular Ecology*, **18**, 4997–5017.

Derome N, Duchesne P, Bernatchez L (2006) Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Molecular Ecology*, **15**, 1239–1249.

Derome N, Bougas B, Rogers SM *et al.* (2008) Pervasive sex-linked effects on transcription regulation as revealed by expression quantitative trait loci mapping in lake whitefish species pairs (*Coregonus* sp., Salmonidae). *Genetics*, **179**, 1903–1917.

Dhar R, Sagesser R, Weikert C, Yuan J, Wagner A (2011) Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *Journal of Evolutionary Biology*, **24**, 1135–1153.

Diz AP, Martinez-Fernandez M, Rolan-Alvarez E (2012) Proteomics in evolutionary ecology: linking the genotype with the phenotype. *Molecular Ecology*, **21**, 1060–1080.

Doerr ED, Dorrough J, Davies MJ, Doerr VA, McIntyre S (2014) Maximizing the value of systematic reviews in ecology when data or resources are limited. *Austral Ecology*. doi: 10.1111/aec.12179.

Dowen RH, Pelizzola M, Schmitz RJ *et al.* (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the USA*, **109**, E2183–E2191.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, **26**, 298–306.

Enard W, Khaitovich P, Klose J *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science*, **296**, 340–343.

Filteau M, Pavey SA, St-Cyr J, Bernatchez L (2013) Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Molecular Biology and Evolution*, **30**, 1384–1396.

Francesconi M, Lehner B (2013) The effects of genetic variation on gene expression dynamics during development. *Nature*, **505**, 208–211.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Gross BL, Kane NC, Lexer C *et al.* (2004) Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *The American Naturalist*, **164**, 145.

Gunter HM, Fan S, Xiong F, Franchini P, Fruciano C, Meyer A (2013) Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. *Molecular Ecology*, **22**, 4516–4531.

Hsieh WP, Chu TM, Wolfinger RD, Gibson G (2003) Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics*, **165**, 747–757.

Hwang WY, Yanfang F, Reyon D *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, **31**, 227–229.

Ilut DC, Coate JE, Luciano AK *et al.* (2012) A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *American Journal of Botany*, **99**, 383–396.

Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L (2010) The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular Ecology*, **19**, 5389–5403.

Johnson JG, Morey JS, Neely MG, Ryan JC, Van Dolah FM (2012) Transcriptome remodeling associated with chronological aging in the dinoflagellate, *Karenia brevis*. *Marine Genomics*, **5**, 15–25.

Kammenga JE, Herman MA, Ouborg N, Johnson L, Breitling R (2007) Microarray challenges in ecology. *Trends in Ecology & Evolution*, **22**, 273–279.

Kendziorski C, Irizarry R, Chen K-S, Haag J, Gould M (2005) On the utility of pooling biological samples in microarray experiments. *Proceedings of the National Academy of Sciences of the USA*, **102**, 4252–4257.

Kilvitis HJ, Alvarez M, Foust CM, Schrey AW, Robertson M, Richards CL (2014) Ecological epigenetics. In: *Ecological Genomics* (eds Landry CR, Aubin-Horth N), pp. 191–210. Springer, New York.

Kobayashi MJ, Takeuchi Y, Kenta T, Kume T, Diway B, Shimizu KK (2013) Mass flowering of the tropical tree *Shorea beccariana* was preceded by expression changes in flowering and drought-responsive genes. *Molecular Ecology*, **22**, 4767–4782.

Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary fisherian life-history evolution in small salmonid populations. *Nature*, **419**, 826–830.

Krasnov A, Koskinen H, Pehkonen P, Rexroad CE, Afanasyev S, Mölsä H (2005) Gene expression in the brain and kidney of rainbow trout in response to handling stress. *BMC Genomics*, **6**, 3.

Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American Journal of Botany*, **99**, 248–256.

Kvist J, Wheat CW, Kallioniemi E, Saastamoinen M, Hanski I, Frilander MJ (2013) Temperature treatments during larval development reveal extensive heritable and plastic variation in gene expression and life history traits. *Molecular Ecology*, **22**, 602–619.

Łabaj PP, Leparc GG, Linggi BE, Markillie LM, Wiley HS, Kreil DP (2011) Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics*, **27**, i383–i391.

Lai Z, Gross BL, Zou Y, Andrews J, Rieseberg LH (2006) Microarray analysis reveals differential gene expression in hybrid sunflower species. *Molecular Ecology*, **15**, 1213–1227.

Landry CR, Aubin-Horth N (2007) Ecological annotation of genes and genomes through ecological genomics. *Molecular Ecology*, **16**, 4419–4421.

Law CW, Chen Y, Shi W, Smyth GK (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, **15**, R29.

Leakey ADB, Ainsworth EA, Bernard SM *et al.* (2009) Gene expression profiling: opening the black box of plant ecosystem responses to global change. *Global Change Biology*, **15**, 1201–1213.

Malone JH, Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology*, **9**, 34.

Meier K, Hansen MM, Normandeau E *et al.* (2014) Local adaptation at the transcriptome level in brown trout: evidence from early life history temperature genomic reaction norms. *PLoS ONE*, **9**, e85171.

de Nadal E, Ammerer G, Posas F (2011) Controlling gene expression in response to stress. *Nature Reviews Genetics*, **12**, 833–845.

Nagano AJ, Sato Y, Mihara M *et al.* (2012) Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. *Cell*, **151**, 1358–1369.

Nicotra AB, Atkin OK, Bonser SP *et al.* (2010) Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, **15**, 684–692.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 332–342.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics*, **32**, 261–266.

Oleksiak MF, Roach JL, Crawford DL (2004) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nature Genetics*, **37**, 67–72.

Page JT, Gingle AR, Udall JA (2013) PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. *G3: Genes | Genomes | Genetics*, **3**, 517–525.

Palumbi SR, Barshis DJ, Traylor-Knowles N, Bay RA (2014) Mechanisms of reef coral resistance to future climate change. *Science*, **344**, 895–898.

Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, **1206**, 110–129.

Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR (2012) What is needed for next-generation ecological and evolutionary genomics? *Trends in Ecology & Evolution*, **27.12**, 673–678.

Pronk TE, van der Veen JW, Ezendam J, Van Loveren H, Pennings JL (2011) Effects of pooling RNA from samples treated with different compounds for determining class specific biomarkers and processes in toxicogenomics. *Toxicology in Vitro*, **25**, 1841–1847.

Rees BB, Andacht T, Skripnikova E, Crawford DL (2011) Population proteomics: quantitative variation within and among populations in cardiac protein expression. *Molecular Biology and Evolution*, **28**, 1271–1279.

Renaut S, Bernatchez L (2011) Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity (Edinburgh)*, **106**, 1003–1011.

Richards CL, Hanzawa Y, Katari MS, Ehrenreich IM, Engelmann KE, Purugganan MD (2009) Perspectives on Ecological and Evolutionary Systems Biology. In: *Plant Systems Biology* (eds Coruzzi G, Gutiérrez R) *Annual Plant Reviews*, **35**, 331–349.

Richards CL, Rosas U, Banta J, Bhambhra N, Purugganan MD (2012) Genome-wide patterns of Arabidopsis gene expression in nature. *PLoS Genetics*, **8**, e1002662.

Roberge C, Guderley H, Bernatchez L (2007) Genomewide identification of genes under directional selection: gene transcription Q(ST) scan in diverging Atlantic salmon subpopulations. *Genetics*, **177**, 1011–1022.

Rosenthal DM, Schwarzbach AE, Donovan LA, Raymond O, Rieseberg LH (2002) Phenotypic differentiation between three ancient hybrid taxa and their parental species. *International Journal of Plant Sciences*, **163**, 387–398.

Scheiner SM, Willig MR, eds (2011) *The Theory of Ecology*. University of Chicago Press, Chicago, Illinois.

Schneider RF, Li Y, Meyer A, Gunter HM (2014) Regulatory gene networks that shape the development of adaptive phenotypic plasticity in a cichlid fish. *Molecular Ecology*, **23**, 4511–4526.

Schraiber JG, Mostovoy Y, Hsu TY, Brem RB (2013) Inferring evolutionary histories of pathway regulation from transcriptional profiling data. *PLoS Computational Biology*, **9**, e1003255.

Shimizu KK, Kudoh H, Kobayashi MJ (2011) Plant sexual reproduction during climate change: gene function in nature studied by ecological and evolutionary systems biology. *Annals of Botany*, **108**, 777–787.

Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, **135**, 367–374.

Trapnell C, Roberts A, Goff L *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**, 562–578.

Ungerer M, Johnson L, Herman M (2007) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*, **100**, 178–183.

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Vera JC, Wheat CW, Fescemyer HW *et al.* (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*, **17**, 1636–1647.

Vijay N, Poelstra JW, Kunstner A, Wolf JB (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular Ecology*, **22**, 620–634.

Webster LM, Paterson S, Mougeot F, Martinez-Padilla J, Piertney SB (2011) Transcriptomic response of red grouse to gastro-intestinal nematode parasites and testosterone: implications for population dynamics. *Molecular Ecology*, **20**, 920–931.

Weigel D (2012) Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant Physiology*, **158**, 2–22.

West-Eberhard MJ (2003) *Developmental Plasticity and Evolution*. Oxford University Press, New York, New York.

Wheat CW, Fescemyer HW, Kvist J *et al.* (2011) Functional genomics of life history variation in a butterfly metapopulation. *Molecular Ecology*, **20**, 1813–1828.

Whitehead A (2012) Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. *Journal of Experimental Biology*, **215**(Pt 6), 884–891.

Whitehead A, Crawford DL (2006a) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the USA*, **103**, 5425–5430.

Whitehead A, Crawford DL (2006b) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197–1211.

Whitehead A, Roach JL, Zhang S, Galvez F (2011) Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proceedings of the National Academy of Sciences of the USA*, **108**, 6193–6198.

Whitehead A, Dubansky B, Bodinier C *et al.* (2012) Genomic and physiological footprint of the Deepwater Horizon oil spill on resident marsh fishes. *Proceedings of the National Academy of Sciences of the USA*, **109**, 20298–20302.

Whitehead A, Zhang S, Roach JL, Galvez F (2013) Common functional targets of adaptive micro- and macro-evolutionary divergence in killifish. *Molecular Ecology*, **22**, 3780–3796.

Whiteley AR, Derôme N, Rogers SM *et al.* (2008) The phenomics and eQTL mapping of brain transcriptomes regulating adaptive divergence in lake whitefish species pairs (*Coregonus* sp.). *Genetics*, **180**, 147–164.

Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. *BMC Evolutionary Biology*, **8**, 282.

Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF (2010) SNP identification, verification, and utility for population genetics in a non-model genus. *BMC Genetics*, **11**, 32.

Wittkopp PJ (2007) Variable gene expression in eukaryotes: a network perspective. *Journal of Experimental Biology*, **210**, 1567–1575.

Wolf JB (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources*, **13**, 559–572.

Wolf JB, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D (2010) Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, **19**(Suppl 1), 162–175.

Wolfinger RD, Gibson G, Wolfinger ED *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, **8**, 625–637.

Wray GA (2013) Genomics and the evolution of phenotypic traits. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 51–72.

Zhang SD, Gant TW (2005) Effect of pooling samples on the efficiency of comparative studies using microarrays. *Bioinformatics*, **21**, 4378–4383.

Zhu J, Zhang B, Smith EN *et al.* (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*, **40**, 854–861.

C.L.R. conceived the original idea for the study. M.A. performed the systematic review and designed the figures. M.A., A.W.S. and C.L.R. wrote the paper.

Data accessibility

Summary of case studies reviewed in Supporting Information.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Case studies reviewed to describe trends in ecological transcriptomic studies in the last 10 years.